Biophysical Journal
Article

Biophysical Society

Enhanced sampling protocol to elucidate fusion peptide opening of SARS-CoV-2 spike protein

Jacob M. Remington,[1] Kyle T. McKay,[1] Jonathon B. Ferrell,[1] Severin T. Schneebeli,[1] and Jianing Li[1,*]
[1]Department of Chemistry, University of Vermont, Burlington, Vermont

ABSTRACT Large-scale conformational transitions in the spike protein S2 domain are required during host-cell infection of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus. Although conventional molecular dynamics simulations have been extensively used to study therapeutic targets of SARS-CoV-2, it is still challenging to gain molecular insight into the key conformational changes because of the size of the spike protein and the long timescale required to capture these transitions. In this work, we have developed an efficient simulation protocol that leverages many short simulations, a dynamic selection algorithm, and Markov state models to interrogate the structural changes of the S2 domain. We discovered that the conformational flexibility of the dynamic region upstream of the fusion peptide in S2 is coupled to the proteolytic cleavage state of the spike protein. These results suggest that opening of the fusion peptide likely occurs on a submicrosecond timescale after cleavage at the S2′ site. Building on the structural and dynamical information gained to date about S2 domain dynamics, we provide proof of principle that a small molecule bound to a seam neighboring the fusion peptide can slow the opening of the fusion peptide, leading to a new inhibition strategy for experiments to confirm. In aggregate, these results will aid the development of drug cocktails to inhibit infections caused by SARS-CoV-2 and other coronaviruses.

SIGNIFICANCE The enveloped viruses of the *Coronaviridae* family have large membrane-anchored spike proteins that are crucial for infection. In particular, the spike protein for SARS-CoV-2 undergoes a large-scale conformational transition from prefusion to postfusion states during host-cell infection that fuses the viral envelope with the host-cell membrane, allowing a passage of viral RNA into the cell. Herein, we interrogate how the proteolytic cleavage of the SARS-CoV-2 spike protein initiates this conformational transition with large-scale molecular simulations. By determining the conformational dynamics of the cleaved spike protein in silico, we demonstrate the targeting of new places on the spike protein with small-molecule drugs to inhibit spike protein opening and thus host-cell infection.

INTRODUCTION

The coronavirus disease 2019 (COVID-19) disease—an infectious disease caused by the coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)—poses a serious public health risk. Despite extensive efforts, of which some have included computational techniques (1,2), specific antiviral medicines to treat the widespread infection are still in critical need. In this work, we have developed an efficient simulation protocol to explore the molecular basis of fusion peptide (FP) opening of the SARS-CoV-2 spike protein (Fig. 1), theoretically indicating a new, to our knowledge, direction to develop antiviral strategies. Distinct from prior and ongoing research to seek therapeutics for the current pandemic, in this study we provide useful computational tools for characterizing spike protein dynamics and an approach to design spike protein inhibitors. As multiple types of the SARS-CoV-2 virus genome have been identified (3), it is essential to design robust antiviral development strategies that can adapt to genomic modifications related to drug resistance and, looking beyond COVID-19, the possibility of additional coronaviruses becoming transmissible to humans (4,5). The large-scale structural changes of the SARS-CoV-2 spike protein (CoVS) that occur during host-cell infection and membrane fusion (Fig. 1) are conserved among other coronaviruses (6). Therefore, seeking therapeutics to restrict these dynamics represents a potential new pathway to inhibit coronavirus infection and will provide critical elements for drug cocktail therapies (7). In this work, we have used theoretical tools to reveal the initial conformational dynamics of CoVS from proteolytically processed models of the prefusion cryo-electron

Submitted November 11, 2020, and accepted for publication May 5, 2021.
*Correspondence: jianing.li@uvm.edu
Editor: Jianhan Chen.
https://doi.org/10.1016/j.bpj.2021.05.022

microscopy (cryo-EM) structure (8) and demonstrated using simulations how therapeutics such as small molecules could inhibit the initial CoVS conformational dynamics and serve as potential fusion inhibitors to block host-cell infection.

In general, the coronavirus spike protein is a large glycoprotein on the viral surface present as a prominent trimer. Each monomer possesses two large domains, S1 and S2, that mainly mediate receptor binding and membrane fusion, respectively (Fig. 1 *A*). During viral infection, spike proteins undergo binding of the S1 domain to receptors of the host cell in step (1). For SARS-CoV-2, host-cell recognition is widely believed to occur through attachment of the receptor binding domain (RBD) of S1 to human ACE2 receptors (11). The RBD of CoVS is quite flexible and can undergo a dynamic opening and closing associated with exposure of the ACE2 binding site (12,13). After attachment of the RBD, proteolytic cleavage at one or both of the S1/S2 (S704–V705) and the S2′ (R815–S816) sites allow dissociation of the S1 domain as step (2). This is based on both the mechanisms of other related viruses, SARS-CoV (14) and Middle East respiratory syndrome-CoV (15), and recent findings about SARS-CoV-2 (16–18). With detachment of the S1 domain, the spike protein undergoes a large-scale multistep conformational transition from prefusion to intermediate and postfusion (steps (3) and (4), respectively) structures, which is expected to be energetically downhill and provide the mechanistic displacement for membrane fusion (6,10). The intermediate opening of the S2 domain in step (3) involves the insertion the FP into the host membrane along with elongation of the heptad repeat (HR) domains HR1/HR2 and connecting regions (Fig. 1 *A*). Finally, refolding of S2 brings the host-membrane-inserted FP into close proximity of the viral enveloped anchored transmembrane domain to pull the two membranes together, resulting in the membrane fusion critical for viral entry (6,19).

Most of the aforementioned steps in SARS-CoV-2 infection are among the targets for antiviral strategies (20,21). Based on its critical role for virus entry, the RBD has become a popular target (22–25). A possible issue, however, is that blocking step (1) with vaccines or small molecules (26–30) can be sensitive to changes in the antigenic sites of the spike protein. Thus, strategies to target other mechanistic steps have drawn attention (31). In particular, the conformational changes of the S2 domain in steps (3) and (4) are required for membrane fusion in viruses across the *Coronaviridae* family (32), which means inhibition of these dynamics with membrane fusion inhibitors could lead to universal and robust antiviral strategies. However, whereas prior results reveal important conformational dynamics associated with step (1) in the SARS-CoV-2 infection cycle, little is known about the conformational changes in CoVS during steps (3) and (4). The results demonstrated herein begin to provide such details that will enable the development of new approaches to inhibit membrane fusion by hindering the insertion of the FP that occurs in step (3).

To design membrane fusion inhibitors, several challenges must be overcome. The first is the choice of which particular conformational state to target. In the past, peptide mimics of the HR1 and HR2 domains have been shown to bind the spike protein in the intermediate state after step (3) and hinder membrane fusion step (4) (33,34). Unfortunately, resistance to these peptides was found in viruses with mutations in the spike protein (35). Furthermore, small-molecule inhibitors to target a transient intermediate state can be difficult to design with high efficacy (36,37). Instead, a promising approach would be targeting the S2 domain before or during (1) or (2) because the inhibitors can bind at any time before the fusion process starts. With design of the correct molecular species, the conformational transition from step (2) to (3), and therefore infection, could be inhibited. However, there is a severe lack of information about how precisely the conformational rearrangement of the S2 domain occurs, and this information is a prerequisite for knowing which parts of the S2 domain to target.

As a powerful tool to reveal mechanistic details of protein structural rearrangements (38–41), molecular dynamics (MD) has been used to help rapidly characterize the SARS-CoV-2 virus in silico by modeling key virus proteins (42) such as the spike protein. Starting from recently determined structures of CoVS, MD simulations have revealed the equilibrium fluctuations of the full-length S1/S2 domain associated with host-cell recognition step (1) including the roles of flexibility of the transmembrane domain in S2 (43), the glycosylation sites in modulating the conformational dynamics of the RBD (44), and the specifics of the protein binding modes of RBD with ACE2 (45,46). Given the system size and large conformational changes, however, guided enhanced sampling is needed to elucidate the dynamics of step (3) and sample relevant conformations. Indeed, to our knowledge, none of the prior efforts mentioned have probed the dynamics of the spike protein during any stage of FP extension into the host-cell membrane. In contrast to prior efforts, we have designed a protocol to capture the dynamics that may occur after proteolytic cleavage and dissociation of S1 by leveraging the anticipated changes of S2 that initiate the insertion of the FP into the host-cell membrane (step (3)). These include the loss of specific intrachain contacts (6) (Figs. 1 and 2). Among the locations expected to diverge, the FP region itself may present a useful target because functional restraints on its sequence make it less mutable and less likely to develop resistance (e.g., see Fig. S1). Ultimately, small molecules could increase these transition state energies by binding and forming bridging interactions to inhibit opening and infection. However, this direction requires the determination of CoVS conformational dynamics during the pre- to postfusion transition. Toward this aim, we present herein extensive molecular simulations of the S2 domain and demonstrate initial steps in this large-scale conformational transition.

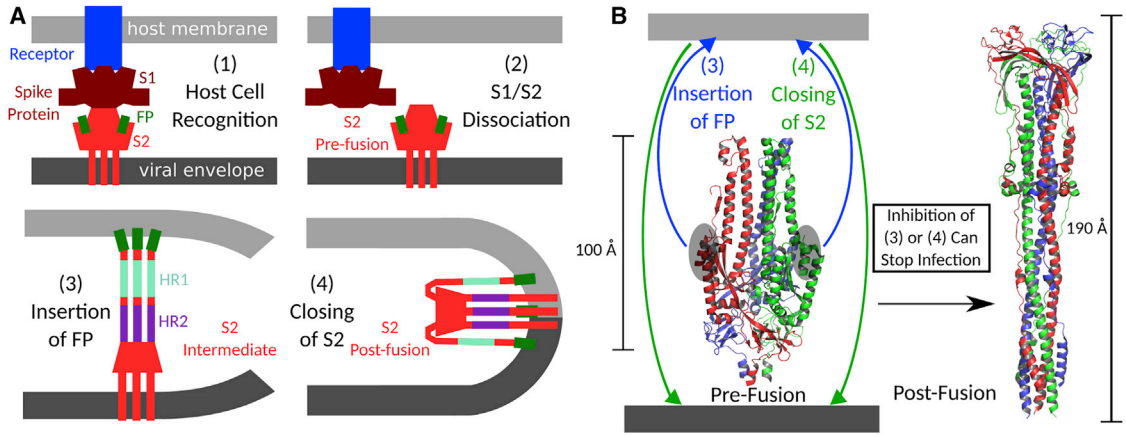

FIGURE 1 (*A*) Overview of *Coronaviridae* viral spike protein mediated membrane fusion. Host-cell recognition (1) initiates cleavage and dissociation of the S1 (residues 1–704) and S2 (residues 705–1273) domains (2) to allow extension of the fusion peptide (FP) into the host membrane (3), which then refolds to drive membrane fusion (4). (*B*) Refolding of the spike protein S2 domain using the SARS-CoV S2 prefusion (PDB: 6VXX) (9) and postfusion (PDB: 6M3W) (10) structures with monomers colored red, green, and blue, respectively. In (*B*), blue and green arrows denote motion of the FP regions (*gray ovals*) that must occur during steps (3) and (4), respectively. To see this figure in color, go online.

MATERIALS AND METHODS

Spike protein sequence alignment

To compare the conservation of amino acids in our binding pockets (Fig. S1), we aligned the CoVS reference sequence (Uniprot: P0DTC2) against reference sequences of other proteins, with the basic local alignment search tool (47). To ensure consistent use of the CoVS residue numbering, additions to the CoVS sequence that were necessary for alignment were skipped. Sequences that were aligned to only subsections of CoVS were ignored. In total, we identified 10 reference sequences to compare, and we present the regions of interest corresponding to the three binding sites (residues 714–718, 721–729, 818–826, 834–853, 914–950, and 914–925).

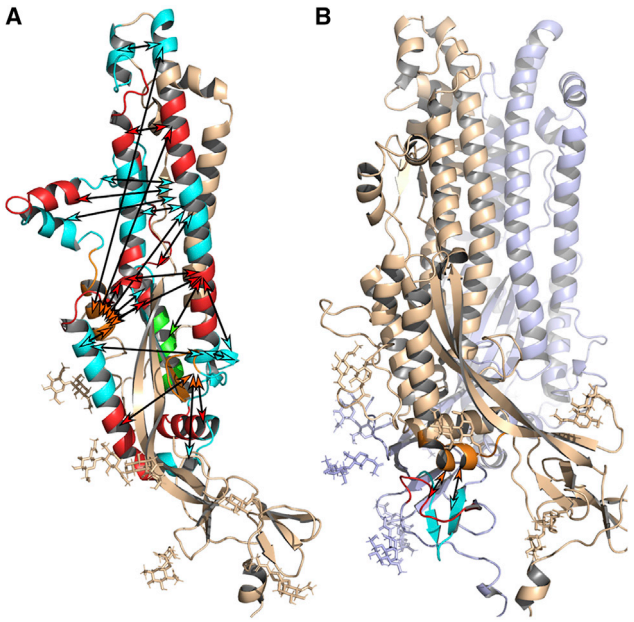

FIGURE 2 (*A*) 23 intramonomer distances between different segments of an individual monomer of the CoVS S2 domain. The arrows are color coded to aid visualization of the different pairs. (*B*) Two intermonomer distances between segments of one monomer and its neighbor. To see this figure in color, go online.

Model preparation and parameters for MD simulations

All-atom models of the SARS-CoV-2 spike glycoprotein homotrimer based on the cryo-EM (Protein Data Bank (PDB): 6VSB) were built using CHARMM-graphical user interface (48). Only the S2 domain (residues: 705–1146), which corresponds to cleavage near the S1/S2 site (8), was included in the S1/S2 cleavage model. Because these models did not include the viral-envelope-anchored domains of the spike protein, a membrane was not included in the simulations. For the S2′ cleavage model, residues 816–1146 were used. Missing residues in the PDB were modeled using galaxyfill (49). These models included four disulfide bonds between residues 739–760, 743–749, 1032–1043, and 1082–1126 for the S1/S2 cleavage model but only the latter two disulfide bonds for the S2′ cleavage model. Additionally, 15 or 9 (for S1/S2 and S2′) glycosylation sites were included at residues 709, 717, 801, 1098, and 1134 of type bDGlcNAc(1 → 4)bDGlcNAc(1 →)asparagine. Each system was placed in a TIP3P water box of dimensions 140 × 140 × 140 $Å^3$ for both S1/S2 and S2′ with 100 mM NaCl plus counterions. This resulted in systems with 257,990 and 257,614 atoms for S1/S2 and S2′, respectively, and provided an initial 30-Å buffer in both directions for the *X* and *Y* dimensions and a 10-Å buffer in each direction of the *Z* dimension to accommodate conformational changes. Because these models did not include the viral-envelope-anchored domains of the spike protein, a membrane was not included in the simulations. This is justified by the lack of involvement of this domain in the initial steps of membrane fusion (steps (1)–(3) in Fig. 1 *A*). However, future simulations targeting more than just the initial steps could utilize the full-length models developed by others (50). The glycoprotein interactions were modeled using the CHARMM-36m additive force field (51). This force field is justified by successful use for other glycoproteins (52) and recently on the CoVS itself (53,54). Ligand force fields were prepared with the CHARMM general force field (55). Hydrogen mass repartitioning was used for these simulations, allowing for a simulation time step of 4 fs (56,57). The systems were first minimized with 2500 steps of steepest decent and 2500 steps of conjugate gradient decent with weak, 1 kcal/(mol $Å^2$) positional restraints on the glycoprotein. A short 125-ps equilibration (with a 1-fs timestep) was then performed to equilibrate the temperature to 310 K with fixed volume and the same positional restraints.

Constant temperature was achieved using a Langevin thermostat with a time constant of 1 ps (58). A 1-ns simulation to equilibrate the pressure was then run with a 4-fs time step and the Monte Carlo barostat. A cutoff for nonbonded interactions of 12 Å was used with a force-based switching at 10 Å. All production simulations were performed with the same parameters as the last equilibration stage using AMBER18 (59) on a single node containing eight NVIDIA Tesla-V100 graphics process units.

nt-PaCS-MD simulations to enhance sampling of initial steps in FP insertion

A variant of the nontargeted parallel cascade selection molecular dynamics (nt-PaCS-MD) simulations (60) was performed to enhance the conformational transitions of the spike protein. In our nt-PaCS-MD simulation methodology, short MD simulations are analyzed on the fly to select rare structures that may represent transition states in a yet unknown transition network. The coordinates of these rare structures are then used to start new short simulations with randomized velocities selected from a Boltzmann distribution. Repeating this process has been shown to enhance conformational sampling, and analyzing the aggregate trajectories with a Markov state model (MSM) allows estimation of the free energy surface (FES) among the conformational states from the equilibrium distribution of the MSM (61). This method falls within the class of simulations where the equilibrium ensemble is estimated by compiling the many microscopic transitions observed from many short simulations that can be run rapidly in parallel (62).

Here, enhanced conformational sampling was achieved along the first four principal components (PCs) (63) of a large set of input intramolecular distances. In this case, 75 intra- and interchain center of mass (COM) distances between sets of 10–11 sequential residues (Fig. 2; Tables S4 and S5) were used for PC determination. These distances were chosen as they are expected to undergo extreme changes during the pre- to postfusion conformational transition (6). PCs can also be used to determine the global modes of proteins by using the aligned coordinates of the protein as input (63), but here we used the PC of the COM distances as a method for collective variable (CV) determination (64–69). Accordingly, at each stage of nt-PaCS-MD, eight 1-ns length simulations (with parameters described in the Model preparation and parameters for MD simulations section above) were performed in parallel. Then, PC analysis was performed on the concatenated 8 ns worth of distances to determine which linear combination of the input distances underwent the most conformational fluctuations, (i.e., had the greatest variance). It was our hypothesis that motion along these loose degrees of freedom could represent initial steps in the CoVS pre- to postfusion transition. Based on this hypothesis, eight frames that maximized or minimized the first four PCs were used to initiate new simulations. This process was repeated 200 times for each system, resulting in 1.6 μs of aggregate simulation time per system. In addition to the nt-PaCS-MD simulations, 400 further 1-ns trajectories starting from structures identified by uniform space clustering of the nt-PaCS-MD trajectories were performed. Structures were saved every 20 ps. Replica simulations were performed using the same initial structures from minimization but different initial velocities for equilibration. For ligand-bound simulations, docking of asiaticoside was performed using Glide (Schrödinger, New York, NY) to determine the optimal binding motif to each chain in the initial structure for MD simulation (see section, Virtual screening and ensemble docking to block the pre- to postfusion transitions). Otherwise, MD simulations were performed identically to those described in the ligand-free simulations.

Analysis of nt-PACS-MD simulations to estimate dynamics of the conformational ensemble

To begin analyzing the simulations, the set of 75 center-of-mass distances used in the nt-PaCS-MD simulations were calculated using Pytraj. Despite the ability of PCs to interrogate the fluctuations within a conformational ensemble, it can be suboptimal for MSM estimation (70,71). To establish a dynamically relevant set of CVs for MSM estimation, the variational approach to Markov processes (VAMP) (72,73) was used within Pyemma (68). In simplified terms, the VAMP establishes a linear combination of input CVs that optimally capture the slowest memoryless processes in the dynamical system (73). Like MSMs, VAMP coordinates also utilize a lag time (VAMP lag time hereafter) for the computation of a time-lagged covariance matrix that captures how fluctuations of a given input variable may be coupled with other input variables at later times. For characterization of the structural heterogeneity explored using nt-PaCS-MD, K-means clustering with eight cluster centers was performed on the first two VAMP coordinates from the nt-PaCS-MD simulations alone with a VAMP lag time of 750 ps (Figs. S2 and S3). For each cluster center, the closest simulation frame, in the Euclidean sense, was found and used for generating the images and root mean-square deviation (RMSD) from initial structure values in Figs. S2 and S3.

For MSM estimation, the first two VAMP coordinates with a VAMP lag time of 980 ps, were chosen as a CV space and clustered using regular space cluster to get 200 cluster centers. Because the structure of the CV space was different for each simulation, the minimal distance parameter in the regular space clustering algorithm was optimized iteratively to achieve 200 cluster centers with a tolerance of one cluster center. Implied timescales (74) from the totality of simulations demonstrate independence of the predicted MSM timescales on MSM lag times larger than 500 ps (Fig. S4). Using a 650-ps MSM lag time, reversible 200-state MSMs were then estimated for each system independently. As with all MSM constructions, this choice of MSM lag time was limited by the length of the shortest simulations used. However, prior MSMs built from short simulations of proteins have demonstrated correct estimation of timescales that are two or more orders of magnitude larger than the shortest simulation length (62,75). Therefore, the 650-ps MSM lag time used herein is likely sufficient to capture the tens to hundreds of nanosecond dynamics of our CoVS cleavage models. The resultant FESs (Figs. 3 and S5) were obtained from Eq. 1, where E_i is the relative free energy of microstate i (in units of kT) and P_i is the probability obtained from the equilibrium distribution of the MSM (i.e., the eigenvector of the transition probability matrix corresponding to an eigenvalue of 1).

$$E_i = -\log(P_i). \tag{1}$$

The two nonstationary eigenvectors with the largest eigenvalues of the MSM characterize the slowest structural dynamics that were resolved with this methodology (Figs. S6–S9). These processes are localized on the low-energy region of the FES and have maximal absolute amplitude near the minimum of the FES. This suggests the nt-PaCS-MD methodology was able to estimate reversible refolding of these structures and not necessarily reversible folding of structures much farther up the FES valleys shown in Figs. 3 and S5. Mean first-passage times were calculated using Pyemma for six transitions between the states corresponding to these three apparent minima in the FES. The helicity of the three states in the S1/S2 model were calculated by counting the number of turns using Pymol's internal secondary structure assignment.

For the ligand-bound simulations, the center-of-mass distance between the FP and the ligand was calculated, and a cutoff of 22 Å was used for quantifying the fraction of bound simulation frames. This cutoff was chosen because it is the first local minimum in the histogram of this CV (Fig. S10). For quantifying restriction of the opening of the FP region in the ligand-bound simulations, Bayesian MSMs were first estimated using the same procedures described above except for only the intramonomer distances (Fig. 2; Tables S4 and S5) being used to allow each chain to be treated independently. Then, states in the MSM that presented a distance between the dynamic region (DR) above the FP and the central helix greater than 30 Å were considered open, and smaller distances were considered closed. This choice was again taken based on the histogram of this distance (Fig. S11). The mean first-passage time (MFPT) from microstates

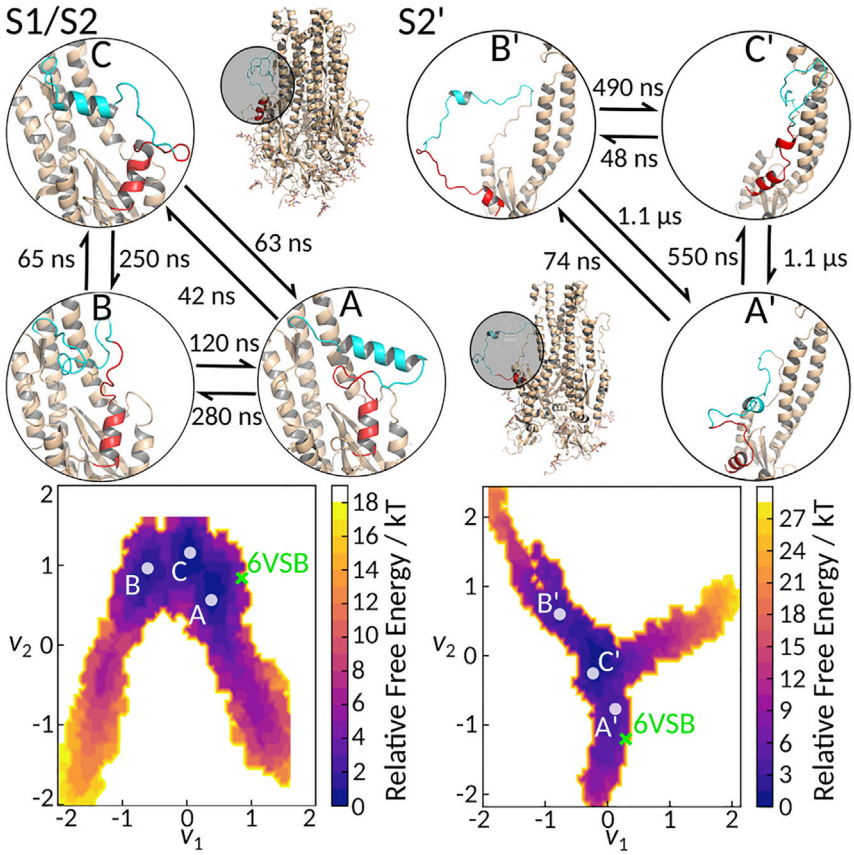

FIGURE 3 MFPTs between states that dominate the two slowest dynamic modes of MSMs (*top*) and FESs (*bottom*) of the S1/S2 (*left*) and S2′ (*right*) cleavage models of CoVS. An overview structure is shown showing the regions of interest (*gray circle*). Structures of states A, B, and C (primes denote S2′) and their positions in the respective VAMP coordinate spaces, v_1 and v_2, are labeled. The fusion peptide is colored red, and the dynamic region is teal. The cryo-EM structure (PDB: 6VSB) (8) is marked with an x to attune the different v_1 and v_2 coordinates for each system. To see this figure in color, go online.

considered closed to microstates considered open was then calculated using 100 samples for error estimation using the Markov chain Monte Carlo algorithm implemented within Pyemma (76). Also included is a calculation of the same MFPT for different choices of cutoff distance (Fig. S12).

Virtual screening and ensemble docking to block the pre- to postfusion transitions

An advantage of targeting the initial opening of the FP is that the more stable conformational ensemble of the nonproteolytically cleaved CoVS containing both the S1 and S2 domains can be used to obtain characteristic structures for the pockets. For the active sites 2 and 3, whose location is described in the following paragraph, the extensive sampling provided by two 10-μs trajectories of the S1 + S2 CoVS complex from the D. E. Shaw research group were obtained and used for the ensemble docking (77). The first of the two provided simulations was initialized from the closed state (PDB: 6VXX), and the second was initialized from a partially open state (PDB: 6VYB). These simulations contained the correct glycosylation and complete protein sequences for all residues in the vicinity of sites 2 and 3. However, the long simulations did not include the aforementioned dynamic loop and were not appropriate for use in the docking to site 1. For active site 1, the initial model for the S2′ simulations described in the Model preparation and parameters for MD simulations was used because it did include the dynamic loop.

Besides the site 1 described in the Results and discussion, two additional sites were chosen for docking to (see Figs. 5 and S13–S15). These sites were chosen by inspecting the differences between the inter-residue contacts of the prefusion and postfusion spike protein conformations. Locations where residues that were close in the prefusion state became far apart in the postfusion state were found. An additional constraint was that these sites be exposed to the solvent in the full spike protein model (containing both S1 and S2 domains). Finally, concave pockets that were large enough to fit a small molecule led to the discovery of sites 1–3. Site 2 is composed of a kink in the α-helix in the HR1 domain below the FP (residues 930–950) that is in close proximity to the β sheet core (residues 721–729) in the prefusion state. These motifs must undergo a "jack-knife" refolding during the opening of the spike protein. A third site, further down this α-helix β sheet zipper, offers a platform for additional anchoring interactions between the base of said helix (residues 914–925) and the β sheet (residues 714–718). A small molecule that could bridge the helix and β sheet in either of these sites could slow or inhibit the spike protein transition. Because the SARS-CoV-2 spike glycoprotein is a homotrimer (with all three sites present on each monomer; Fig. 5) and the following PC analysis is time independent, each monomer of each trajectory was considered its own simulation for the conformational analysis. In this way, 60 μs of simulation data were analyzed per active sites 2 and 3 to determine characteristic structures of the targeted pockets.

Visual MD (78) was used to align the S2 helical domain of the protein to the initial coordinates (crystal structure from PDB) to reduce noise due to translational movement. The *XYZ* coordinates of the residues comprising active site 2 (residues 721–729, residues 930–950, and residue 1028) and active site 3 (residues 714–718, residues 914–925, residues 1108–1113, and residues 1148–1149) were collected and parsed using Tcl and python. A dimension reduction (PC analysis) was used to determine which residue movements contributed most to variations in the overall conformations of each active site. For each active site, the first 10 dimensions (accounting for ~90% of the cumulative variance) were clustered using a K-means clustering algorithm (Fig. S16).

Each cluster center represents a unique active-site conformation in the overall conformational ensemble. As these cluster centers do not represent frames but instead are points in the principal component space, an N-dimensional RMSD was used to find representative frames closest in value to these cluster centers. These frames were then prepared using Maestro protein preparation utility, and the docking grids were prepared

using Glide (Schrödinger release 2018-2). The receptor grids representing the conformational ensemble were screened against the Food and Drug Administration (FDA)-approved drug database that included 2682 unique molecules acquired in the structure-data file format (79) and prepared using LigPrep software (release 2018-2). Two levels of docking precision were used. First, high-throughput virtual screening (HTVS) precision was used to quickly eliminate ligands with little or no binding affinity. The top 5% of ligands were then screened again using the extra precision (XP) level of precision to give a more accurate binding affinity. The top 40 ligands were kept. Of these ligands, all ligands not in the top 40 of the other conformation were cross docked against the other receptors to get data points for all ligand-protein complex combinations. The final docking score was calculated as a Boltzmann weighted average of all unique ligand-protein complex Glide XP scores (Tables S1–S3).

RESULTS AND DISCUSSION

As a first step to determine the conformational dynamics of CoVS that initiate the pre- to postfusion transition (steps (3)–(4) in Fig. 1 *A*), we performed extensive MD simulations of the S2 domain. Specifically, we focused on two all-atom models of CoVS with proteolytic cleavages of all three monomers at the S1/S2 or S2′ sites, respectively (the S1/S2 cleavage model contains residues 705–1146, and the S2′ cleavage model contains residues 816–1146) (9), corresponding to a CoVS state just after S1 dissociation and ready to begin S2 opening and refolding (steps (3)–(4) in Fig. 1). Although there are many more possible combinations of initial models because of the presence of three monomers with two cleavage sites each, we chose our models to represent two extreme cases in which all the monomers are cleaved at either the S1/S2 site or the S2′ site. These simulations, with two replicas per system totaling 8 μs, were conducted in the spirit of nt-PaCS-MD (60,61). Accordingly, many short MD simulations were analyzed on the fly to select rare conformations on the FES corresponding to the initial unfolding of the S2 domain (Fig. 2). Repetition of this process resulted in enhanced sampling of S2 domain conformations in the aggregate trajectories. Using the results of an MSM, the FES and kinetics among the low-energy conformational states were estimated (61,74,80).

Strikingly, we observed significantly distinct conformational fluctuations of CoVS in these two constructs with S1/S2 cleavage and S2′ cleavage, which may indicate different ability to undergo opening. The first major difference is in the conformational heterogeneity of CoVS (Figs. S2 and S3). The RMSD relative to the initial structure for 10 representative structures is much larger for the S2′ cleavage model (5.4–10.9 Å) than for the S1/S2 cleavage model (3.5–5.8 Å). This suggests that the loss of residues 705–815 between the S1/S2 sites and the S2′ site has a major impact on the flexibility of CoVS. Second, the S2′ simulations displayed more conformations with extended FP than the S1/S2 simulations, suggesting that this model is more primed than the S1/S2 cleavage model for FP insertion and the subsequent large-scale conformational transitions that initiate the pre- to postfusion transition.

Then, we constructed reversible MSMs to more concretely examine the conformational transitions and energetics. For each system, the FESs along the two slowest VAMP coordinates (72,73), v_1 and v_2, were obtained (see additional details in the Materials and methods). The FESs for both systems (Figs. 3 and S5) are characterized by broad regions of low relative free energy, less than 4 kT, inhabiting two (for S1/S2) or three (for S2′) long valleys. In the low-energy regions, three local minima are apparent in each system denoted as states A, B, and C for S1/S2 (Fig. 3) and A′, B′, and C′ for S2′ (Fig. 3). In a self-consistent manner, these three states also correspond to states with maximal involvement in the first two slowest dynamic modes (Figs. S6 and S7). Such correspondence between the slowest modes of the MSM and the minima in the FES supports a converged picture of the dynamics near the initial structure because they represent reversible folding of these states. Similar features of the FES and the MSM were observed for replica simulations of each system, in which the three dominant processes observed in the first two MSM eigenvectors reside in the low-energy regions of the FES. On the other hand, the two replicas did not result in MSMs with the same relaxation timescales (Figs. S6–S9). The relaxation timescales are not used directly in the following analysis, which instead are focused on the consistent features shared between the replica simulations. It should be noted that the VAMP coordinates v_1 and v_2 in the replica simulations are not necessarily the same in two replicas, which can also lead to differences in the precise features resolved. However, despite this drawback, the predominant states share significant qualitative resemblance with low RMSDs (3.2–4.2 Å) between replicas for the S1/S2 cleavage model and slightly higher RMSDs (5.3–7.3 Å) for the S2′ cleavage model. Based on these observations, a simple three-state model was built by estimating the MFPT for the six transitions between the three states (Figs. 3 and S5).

Focusing on the S1/S2 cleavage model first, the three states correspond to the different degrees of displacement and helicity of the region just after the FP with helicity ranked roughly from 0 to 2.5 to 3.5 helical turns for states B, C, and A. Based on the MFPT from B to A and C (120 and 65 ns) compared with the reverse processes (280 and 250 ns) and the lower relative energy ($\sim$1 kT) of the helical A/C regions, the more helical states A and C are favored. For the replica simulation of the S1/S2 cleavage model, the nonhelical conformational of the DR (state B) was not resolved by the MSM (Fig. S5), but the more helical state still appeared more favorable. Generally, and more importantly, however, the dynamics of this region in the S1/S2 system are confined and do not correspond to extensions of the FP region observed in the nt-PaCS-MD (Figs. S2 and S3). Such extended structures reside in the higher-energy regions further up the valleys of the FES and would likely require more sampling to more accurately measure their kinetics.

In contrast to the S1/S2 cleavage model, the S2′ cleavage model captured large-scale displacements of residues including the FP (Fig. 3, state B′). Similar extended structures were also observed in the replica simulation (Fig. S5, state B′), although the flexibility of the extended conformations, and the S2′ model in general, resulted in larger RMSD between the replica structures (5.3–7.3 Å). In fact, the low-energy state B′ is predicted to be dynamically favored because the MFPTs into state B′ are at least an order of magnitude faster than transitions out of state B (74 ns compared to 1.1 μs for B′/A′ and 48 and 490 ns for B′/C′). Given this finding, the proximity of state A to the initial cryo-EM structure, and the obvious extension of the FP in state B, state B may represent an initial stage in the pre- to postfusion transition of CoVS. Overall, our model suggests that cleavage at S2′ facilitates release of the FP sequence to extend outward form the core of CoVS on a submicrosecond timescale. These computational results provide an atomistic understanding of prior reports that cleavage at the S2′ site is required for SARS-CoV (14) and SARS-CoV2 (81) virus infectivity.

The dynamic opening of the FP observed in the MD simulations presents a target for small-molecule inhibitors. Specifically, the observed dynamics are in general agreement with the proposed mechanism for spike protein action (initiation of step 3 in Fig. 1 *A*). This suggests that hindering this opening event with small molecules bound to specific sites on CoVS could slow down the infection process. The opening occurs at the interface between the FP and the HR1 domains. Toward achieving this task, HTVS was used to find molecules that could inhibit the opening of the FP with molecular docking of the FDA-approved drug database (79) to the seam presented between the FP and a kink in the HR1 helix (respectively, in site 1; Figs. 5 and S13). This site lays above the rigid core β sheet of CoVS (in this case, the loop composed of residues A1056-G1059), between the helical part of the FP (residues I818–V826), the DR above the FP (residues 834–853), and the HR1 kink (residues 940–946). 20 molecules with binding affinities ranging from -7 to -10 kcal/mol were identified (Table S1). In particular, asiaticoside (CAS 16830-15-2), a biologically active compound with well-studied wound healing properties (82), appeared to form substantial contacts between HR1 and the FP (*complex structure* in Fig. 4).

To test the ability of asiaticoside to inhibit the FP opening, the same simulation methodologies were performed on an S2′ cleavage model with an asiaticoside molecule bound to each chain. In total, the FP region remained closed during 97% of the time the ligand was bound. We note that because the enhanced sampling was designed to open S2′ and disrupt the binding pocket, the ligand only remained bound to the FP region for 52% of the enhanced sampling simulations. To better quantify asiaticoside's ability to restrict opening, the MFPT between the set of closed and open states in the MSM was calculated and compared to

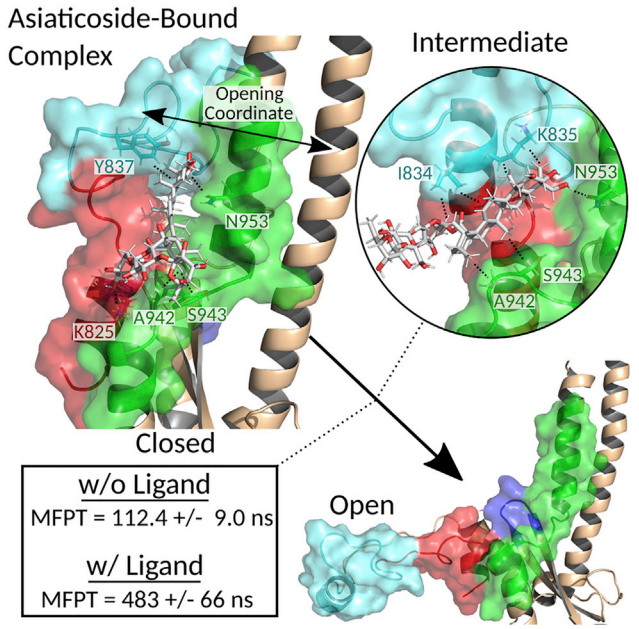

FIGURE 4 Inhibition of FP opening by the ligand asiaticoside as judged by MFPTs from the closed to open states. The ligand (*white* and *red licorice*) forms bridging contacts between the HR1 (*green*), the FP (*red*), and the DR (*teal*), preventing opening and the exposure of a loop on the core β-sheet (*blue*). Error estimation was performed using the Markov chain Monte Carlo algorithm implemented within Pyemma (76). To see this figure in color, go online.

the apo-S2′ simulation (Fig. 4). This demonstrated an over fourfold increase in the timescale for this step in S2′ opening from 112 $\pm$ 9 to 483 $\pm$ 66 ns. A slower MFPT was found for the asiaticoside-bound simulations across a range of cutoff distances to distinguish closed and open states, although the slowdown did depend on the chosen cutoff (Fig. S12). Investigating an intermediate structure with a distance between the open and closed states, we observed the asiaticoside ligand forming bridging contacts between I834 and K835 of DR and A942, S943, and N953 of HR1. Furthermore, for the apo-S2′ simulation, the fraction of frames with the closed FP was 0.70, whereas for the asiaticoside-bound frames, it was 0.81. These observations, taken together, provide proof of principle that small molecules may be able to inhibit the opening of CoVS.

However, this simulation methodology could only observe the initial steps in opening of the spike protein, and many other steps are necessary for refolding during membrane fusion. We expect at least two other sites (Fig. 5; see Materials and methods for details) could also provide hit compounds, further unlocking the possibility for drug cocktails with binding to more than one site at a time. Blast alignment of the CoVS sequence demonstrate that these sites have low mutability supporting their druggability (Fig. S1). To facilitate the search for such molecules, HTVS and ensemble molecular docking were employed using 20 μs of simulation time provided for free to the scientific community from the D. E.

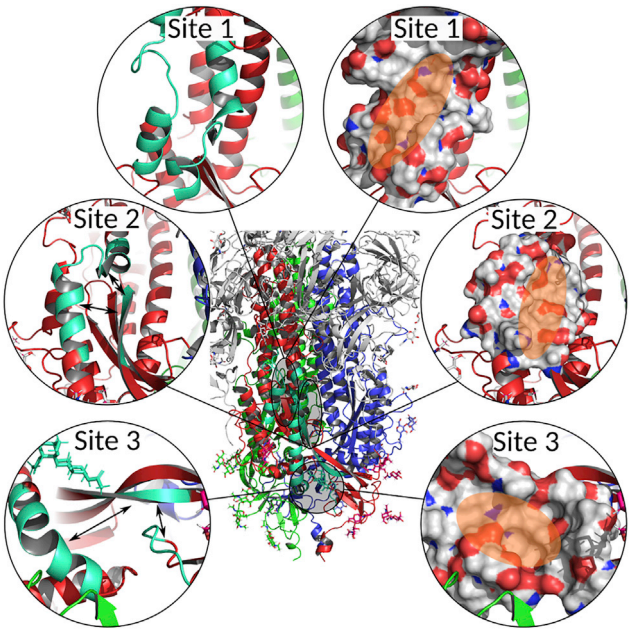

FIGURE 5 Three sites on SARS-CoV-2 spike protein that have intrachain contacts (*arrows* in *left*) that small molecules could bridge. On the right, a surface representation is shown to demonstrate the concave pockets formed at these sites that could facilitate strong binding. Ligands bound to site 1 were used in the main text for ligand-bound simulations. Docking results for all three sites are shown in Tables S1–S3. The monomers of the S2 domain of the spike protein are shown in red, blue, and green, and the highlighted residues in each site are teal. The S1 domain of the spike protein is shown in white. To see this figure in color, go online.

Shaw research group (77). Ligands were taken from the FDA-approved drug database and screened against 10 conformations of each active site (Fig. S16). For both sites, binding affinities ranging from −7 to −20 kcal/mol as well as ligands with substantial bridging interactions were identified (examples of docked motifs in Figs. S14 and S15 and specific residue-ligand interactions with two-dimensional chemical structures for sites 1–3 in Fig. S17). Running ligand-bound simulations for opening of these sites will require increases in simulation power and/or more targeted enhanced sampling methodologies. However, based on the findings at the site near the FP (Fig. 4), where the formation of a few bridging interactions between adjacent domains resulted in a slowdown in FP opening, it is likely that the ligands that form even more extensive bridging interactions at sites 2 and 3 could also restrict the protein dynamics associated with membrane fusion.

CONCLUSIONS

In summary, we investigated the dynamics of the CoVS during the early stages of host-cell infection with state-of-the-art enhanced sampling techniques. By using a collection of specific interchain distances that increase during the FP opening of CoVS, we achieve rapid sampling of this process. This demonstrates an alternative way to select the many initial structures required when using many short MD runs to explore configuration space (62), by selecting conformations that maximize the fluctuations of the structural ensemble. This approach exploits the properties of PC analysis in that the PCs with largest eigenvalues represent directions in conformation space where the structure is most flexible. In our specific application, the results of the simulations provide molecular-level evidence delineating the role of specific cleavage sites on the CoVS. In particular, we demonstrate that proteolytic cleavage at the S2′ site (R815) can facilitate opening of the FP region more so than cleavage at just the S1/S2 site. The loss of the S2 residues between 705 and 815 in the S2′ cleavage site model appears to destabilize the dynamic loop above the FP, and the FP itself, to initiate its eventual insertion into the bilayer. Based on our mechanistic findings and the anticipated changes in the CoVS during refolding, we propose three binding sites for potential drug candidates to inhibit SARS-COV-2 infection. Multiple FDA-approved compounds with high-binding affinities to these sites were found computationally in this work, and in silico predictions demonstrate that at least one is able to slow down the opening of the spike protein. Not only do these results suggest new directions to design drugs that combat COVID-19, but our methodology can also be readily adapted to other spike protein containing viruses to allow the development of a new class of antivirals targeting the FP region and the conformational transitions that occurring during the pre- to postfusion transitions. Notably, among the recently discovered virus variants (United Kingdom (83), South Africa (84), Japan (85), and United States (86)), only the T716I mutation in the United Kingdom and United States variants falls within one of the three binding sites (site 3) we suggest herein. Thus, our proposed approach to design fusion inhibitors necessitates additional research and experimental validation. Even if just weak activity is verified experimentally for the compounds determined herein, these compounds target binding sites and stages of the SARS-COV-2 infection cycle different from currently known drugs, which will enable effective cocktail therapies that simultaneously target multiple aspects of *Coronaviradae* infection.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2021.05.022.

AUTHOR CONTRIBUTIONS

The manuscript was written through contributions of all authors. J.M.R. and J.B.F. designed the MD simulation protocol. J.M.R. and K.T.M. found the binding pockets. K.T.M. conducted all HTVS and ensemble docking. J.B.F. performed the sequence analysis. J.M.R. selected the CVs, implemented, and analyzed the simulations. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Professors Yvonne Janssen-Heininger and Jason Botten at the University of Vermont for helpful discussions. J.M.R. and K.T.M. were supported by a National Institutes of Health award (R01GM129431 to J.L.). J.L. and J.B.F. were partially supported by a National Science Foundation award (CHE-1945394 to J.L.). S.T.S. was supported by the U.S. Army Research Office (grant 71015-CH-YIP). Computations were performed on the Vermont Advanced Computing Core supported in part by National Science Foundation award no. OAC-1827314.

REFERENCES

1. Morgnanesi, D., E. J. Heinrichs, …, J. L. Kulp, III. 2015. A computational chemistry perspective on the current status and future direction of hepatitis B antiviral drug discovery. *Antiviral Res.* 123:204–215.
2. Mirza, M. U., M. Vanmeert, …, M. Idrees. 2019. Perspectives towards antiviral drug discovery against Ebola virus. *J. Med. Virol.* 91:2029–2048.
3. Forster, P., L. Forster, …, M. Forster. 2020. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc. Natl. Acad. Sci. USA.* 117:9241–9243.
4. Li, W., Z. Shi, …, L.-F. Wang. 2005. Bats are natural reservoirs of SARS-like coronaviruses. *Science.* 310:676–679.
5. Woo, P. C., S. K. Lau, …, K. Y. Yuen. 2009. Coronavirus diversity, phylogeny and interspecies jumping. *Exp. Biol. Med. (Maywood).* 234:1117–1127.
6. Walls, A. C., M. A. Tortorici, …, D. Veesler. 2017. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. USA.* 114:11157–11162.
7. Hung, I. F.-N., K.-C. Lung, …, K.-Y. Yuen. 2020. Triple combination of interferon beta-1b, lopinavir-ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. *Lancet.* 395:1695–1704.
8. Wrapp, D., N. Wang, …, J. S. McLellan. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science.* 367:1260–1263.
9. Walls, A. C., Y.-J. Park, …, D. Veesler. 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell.* 181:281–292.e6.
10. Fan, X., D. Cao, …, X. Zhang. 2020. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nat. Commun.* 11:3618.
11. Ou, X., Y. Liu, …, Z. Qian. 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* 11:1620.
12. Gur, M., E. Taka, …, M. Golcuk. 2020. Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. *J. Chem. Phys.* 153:075101.
13. Xiong, X., K. Qu, …, J. A. G. Briggs; CITIID-NIHR COVID-19 BioResource Collaboration. 2020. A thermostable, closed SARS-CoV-2 spike protein trimer. *Nat. Struct. Mol. Biol.* 27:934–941.
14. Belouzard, S., V. C. Chu, and G. R. Whittaker. 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. USA.* 106:5871–5876.
15. Kleine-Weber, H., M. T. Elzayat, …, S. Pöhlmann. 2018. Functional analysis of potential cleavage sites in the MERS-coronavirus spike protein. *Sci. Rep.* 8:16597.
16. Hoffmann, M., H. Kleine-Weber, and S. Pöhlmann. 2020. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol. Cell.* 78:779–784.e5.
17. Hoffmann, M., H. Kleine-Weber, …, S. Pöhlmann. 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* 181:271–280.e8.
18. Vankadari, N. 2020. Structure of furin protease binding to SARS-CoV-2 spike glycoprotein and implications for potential targets and virulence. *J. Phys. Chem. Lett.* 11:6655–6663.
19. Rey, F. A., and S.-M. Lok. 2018. Common features of enveloped viruses and implications for immunogen design for next-generation vaccines. *Cell.* 172:1319–1334.
20. Jeong, G. U., H. Song, …, Y.-C. Kwon. 2020. Therapeutic strategies against COVID-19 and structural characterization of SARS-CoV-2: a review. *Front. Microbiol.* 11:1723.
21. Su, H., F. Zhou, …, H. Su. 2021. Molecular insights into small-molecule drug discovery for SARS-CoV-2. *Angew. Chem. Int. Ed. Engl.* 60:9789–9802.
22. Tai, W., L. He, …, L. Du. 2020. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell. Mol. Immunol.* 17:613–620.
23. Tai, W., X. Zhang, …, L. Du. 2020. Identification of SARS-CoV RBD-targeting monoclonal antibodies with cross-reactive or neutralizing activity against SARS-CoV-2. *Antiviral Res.* 179:104820.
24. Choudhary, S., Y. S. Malik, and S. Tomar. 2020. Identification of SARS-CoV-2 cell entry inhibitors by drug repurposing using in silico structure-based virtual screening approach. *Front. Immunol.* 11:1664.
25. Cao, L., I. Goreshnik, …, D. Baker. 2020. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science.* 370:426–431.
26. Adedeji, A. O., W. Severson, …, S. G. Sarafianos. 2013. Novel inhibitors of severe acute respiratory syndrome coronavirus entry that act by three distinct mechanisms. *J. Virol.* 87:8017–8028.
27. Zhang, H., J. M. Penninger, …, A. S. Slutsky. 2020. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med.* 46:586–590.
28. Chen, Y., Y. Guo, …, Z. J. Zhao. 2020. Structure analysis of the receptor binding of 2019-nCoV. *Biochem. Biophys. Res. Commun.* 525:135–140.
29. Han, Y., and P. Král. 2020. Computational design of ACE2-based peptide inhibitors of SARS-CoV-2. *ACS Nano.* 14:5143–5147.
30. Hasan, A., B. A. Paray, …, M. Falahati. 2021. A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin. *J. Biomol. Struct. Dyn.* 39:3025–3033, Published online April 22, 2020.
31. Xia, S., M. Liu, …, L. Lu. 2020. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* 30:343–355.
32. Heald-Sargent, T., and T. Gallagher. 2012. Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses.* 4:557–580.
33. Bosch, B. J., B. E. E. Martina, …, P. J. M. Rottier. 2004. Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *Proc. Natl. Acad. Sci. USA.* 101:8455–8460.
34. Zhu, J., G. Xiao, …, G. F. Gao. 2004. Following the rule: formation of the 6-helix bundle of the fusion core from severe acute respiratory syndrome coronavirus spike protein and identification of potent peptide inhibitors. *Biochem. Biophys. Res. Commun.* 319:283–288.
35. Bosch, B. J., J. W. A. Rossen, …, P. J. M. Rottier. 2008. Coronavirus escape from heptad repeat 2 (HR2)-derived peptide entry inhibition as a result of mutations in the HR1 domain of the spike fusion protein. *J. Virol.* 82:2580–2585.
36. Schramm, V. L. 2018. Enzymatic transition states and drug design. *Chem. Rev.* 118:11194–11258.
37. Vajda, S., D. Beglov, …, A. Whitty. 2018. Cryptic binding sites on proteins: definition, detection, and druggability. *Curr. Opin. Chem. Biol.* 44:1–8.
38. Chodera, J. D., and F. Noé. 2014. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* 25:135–144.

39. Guo, J., and H.-X. Zhou. 2016. Protein allostery and conformational dynamics. *Chem. Rev.* 116:6503–6515.

40. Li, J., A. L. Jonsson, …, G. A. Voth. 2013. Ligand-dependent activation and deactivation of the human adenosine A(2A) receptor. *J. Am. Chem. Soc.* 135:8749–8759.

41. Liao, C., V. May, and J. Li. 2019. PAC1 receptors: shapeshifters in motion. *J. Mol. Neurosci.* 68:331–339.

42. Wang, Y., M. Liu, and J. Gao. 2020. Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. *Proc. Natl. Acad. Sci. USA.* 117:13967–13974.

43. Turoňová, B., M. Sikora, …, M. Beck. 2020. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science.* 370:203–208.

44. Casalino, L., Z. Gaieb, …, R. E. Amaro. 2020. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent. Sci.* 6:1722–1734.

45. Ali, A., and R. Vijayan. 2020. Dynamics of the ACE2-SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. *Sci. Rep.* 10:14214.

46. Roy, S., A. Jaiswar, and R. Sarkar. 2020. Dynamic asymmetry exposes 2019-nCoV prefusion spike. *J. Phys. Chem. Lett.* 11:7021–7027.

47. Altschul, S. F., W. Gish, …, D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.

48. Jo, S., T. Kim, …, W. Im. 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29:1859–1865.

49. Coutsias, E. A., C. Seok, …, K. A. Dill. 2004. A kinematic view of loop closure. *J. Comput. Chem.* 25:510–528.

50. Woo, H., S.-J. Park, …, W. Im. 2020. Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane. *J. Phys. Chem. B.* 124:7128–7137.

51. Huang, J., S. Rauscher, …, A. D. MacKerell, Jr. 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods.* 14:71–73.

52. Lilkova, E., P. Petkov, …, L. Litov. 2019. Molecular modeling of the effects of glycosylation on the structure and dynamics of human interferon-gamma. *J. Mol. Model.* 25:127.

53. McAuley, A. J., M. J. Kuiper, …, S. S. Vasan. 2020. Experimental and in silico evidence suggests vaccines are unlikely to be affected by D614G mutation in SARS-CoV-2 spike protein. *NPJ Vaccines.* 5:96.

54. Pandey, P., J. S. Rane, …, S. Ray. 2020. Targeting SARS-CoV-2 spike protein of COVID-19 with naturally occurring phytochemicals: an in silico study for drug development. *J. Biomol. Struct. Dyn* 1–11, Published online July 22, 2020.

55. Vanommeslaeghe, K., E. Hatcher, …, A. D. Mackerell, Jr. 2010. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* 31:671–690.

56. Hopkins, C. W., S. Le Grand, …, A. E. Roitberg. 2015. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* 11:1864–1874.

57. Balusek, C., H. Hwang, …, J. C. Gumbart. 2019. Accelerating membrane simulations with hydrogen mass repartitioning. *J. Chem. Theory Comput.* 15:4673–4686.

58. Loncharich, R. J., B. R. Brooks, and R. W. Pastor. 1992. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N′-methylamide. *Biopolymers.* 32:523–535.

59. Lee, T.-S., D. S. Cerutti, …, D. M. York. 2018. GPU-accelerated molecular dynamics and free energy methods in Amber18: performance enhancements and new features. *J. Chem. Inf. Model.* 58:2043–2050.

60. Harada, R., V. Sladek, and Y. Shigeta. 2019. Nontargeted parallel cascade selection molecular dynamics using time-localized prediction of conformational transitions in protein dynamics. *J. Chem. Theory Comput.* 15:5144–5153.

61. Harada, R., and A. Kitao. 2013. Parallel cascade selection molecular dynamics (PaCS-MD) to generate conformational transition pathway. *J. Chem. Phys.* 139:035103.

62. Noé, F., C. Schütte, …, T. R. Weikl. 2009. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. USA.* 106:19011–19016.

63. David, C. C., and D. J. Jacobs. 2014. Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol. Biol.* 1084:193–226.

64. Sittel, F., and G. Stock. 2018. Perspective: identification of collective variables and metastable states of protein dynamics. *J. Chem. Phys.* 149:150901.

65. Lätzer, J., T. Shen, and P. G. Wolynes. 2008. Conformational switching upon phosphorylation: a predictive framework based on energy landscape principles. *Biochemistry.* 47:2110–2122.

66. Hori, N., G. Chikenji, …, S. Takada. 2009. Folding energy landscape and network dynamics of small globular proteins. *Proc. Natl. Acad. Sci. USA.* 106:73–78.

67. Allen, L. R., S. V. Krivov, and E. Paci. 2009. Analysis of the free-energy surface of proteins from reversible folding simulations. *PLoS Comput. Biol.* 5:e1000428.

68. Scherer, M. K., B. Trendelkamp-Schroer, …, F. Noé. 2015. PyEMMA 2: a software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* 11:5525–5542.

69. Kalgin, I. V., A. Caflisch, …, M. Karplus. 2013. New insights into the folding of a β-sheet miniprotein in a reduced space of collective hydrogen bond variables: application to a hydrodynamic analysis of the folding flow. *J. Phys. Chem. B.* 117:6092–6105.

70. Schwantes, C. R., and V. S. Pande. 2013. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* 9:2000–2009.

71. Pérez-Hernández, G., F. Paul, …, F. Noé. 2013. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* 139:015102.

72. Mardt, A., L. Pasquali, …, F. Noe. 2018. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* 9:11.

73. Wu, H., and F. Noé. 2020. Variational approach for learning Markov processes from time series data. *J. Nonlinear Sci.* 30:23–66.

74. Pande, V. S., K. Beauchamp, and G. R. Bowman. 2010. Everything you wanted to know about Markov state models but were afraid to ask. *Methods.* 52:99–105.

75. Sun, X., S. Singh, …, G. R. Bowman. 2018. Simulation of spontaneous G protein activation reveals a new intermediate driving GDP unbinding. *eLife.* 7:e38465.

76. Trendelkamp-Schroer, B., H. Wu, …, F. Noé. 2015. Estimation and uncertainty of reversible Markov models. *J. Chem. Phys.* 143:174101.

77. D. E. Shaw Research. 2020. Molecular dynamics simulations related to SARS-CoV-2 https://www.deshawresearch.com/downloads/download_trajectory_sarscov2.cgi/.

78. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.

79. https://www.selleckchem.com/screening/fda-approved-drug-library.html.

80. Husic, B. E., and V. S. Pande. 2018. Markov state models: from an art to a science. *J. Am. Chem. Soc.* 140:2386–2396.

81. Xia, S., Q. Lan, …, S. Jiang. 2020. The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin. *Signal Transduct. Target. Ther.* 5:92.

82. Lee, J. H., H. L. Kim, …, J. C. Park. 2012. Asiaticoside enhances normal human skin cell migration, attachment and growth in vitro wound healing model. *Phytomedicine.* 19:1223–1227.

83. Rambaut, A., N. Loman, …, E. Volz; on behalf of COVID-19 Genomics Consortium UK (CoG-UK). 2020. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK

defined by a novel set of spike mutations https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563.

84. Tegally, H., E. Wilkinson, …, T. de Oliveira. 2020. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* https://doi.org/10.1101/2020.12.21.20248640.

85. Fujino, T., N. Hidetoshi, K. Satoshi…, 2021. Novel SARS-CoV-2 variant identified in travelers from Brazil to Japan. *Emerg. Infect. Dis.* 27:1243–1245. https://doi.org/10.3201/eid2704.210138.

86. Tu, H., M. R. Avenarius, …, D. Jones. 2021. Distinct patterns of emergence of SARS-CoV-2 spike variants including N501Y in clinical samples in columbus Ohio. *bioRxiv* https://doi.org/10.1101/2021.01.12.426407.